Computers Math. Applic. Vol. 23, No. 6–9, pp. 419–440, 1992
Printed in Great Britain.

0097-4943/92 $5.00 + 0.00

A DESIGN APPROACH FOR CONSTRUCTING ENGINEERING SCENARIO MAPS

ROB RUCKER AND TARIQ A. ALDOWAISAN
Department of Industrial and Management Systems Engineering and
Center for Computer Integrated Manufacturing
Arizona State University, Tempe, Arizona 85287, U.S.A.

Abstract—We present a cognitively based design approach for the staged construction of a high level linguistic-visual map useful for engineering scenario analysis. This map, which we call a Three Dimensional Conceptual Thematic Map (3D-CT Map), is a linkage of a 3-Dimensional geometric Map (3D-Map) and a semantic net which we have called a Conceptual Thematic Map (CT-Map). The 3D-CT Map is an attempt to specify what is in the environment, where it is, and what is happening to it. The CT-Map component is derived by combining information from two explicit linguistic levels, syntax and semantics. It consists of recursively nested semantic structures linked by thematic roles. The 3D-Map component is derived by combining the information from two explicit visual levels, the 2 1/2 D sketches (which correspond to standard engineering drawings) and three dimensional shape models. The result of these visual combinations is the 3D-Map which shows the object centered three dimensional geometric component of the scenario. Annotating the 3D-Map with selected cross references to the CT-Map constitutes the 3D-CT Map.

The theoretical bases of the 3D-CT Map rest on the combined insights of contemporary vision and linguists researchers, principally Ray Jackendoff and David Marr. These linguistic and vision insights are founded on a representationalist view of human understanding and action that includes the formal recognition, analysis, and constructive representation of autonomous levels of mental information structures. Each level of representation has its own set of primitives, well-formedness rules, and links to other levels via inter and intra level correspondence rules. At an even more fundamental level, these insights are in consonance with a view of the human mind/brain as a biological information processor.

We illustrate this cognitive design approach by constructing a 3D-CT Map from a scenario drawn from the spatial domain of Numerical Control Part Programming. The inputs to the derivation consist of engineering drawings, a natural-language scenario description of the procedure to be carried out, and the experience of the part programmer. The outputs from the scenario analysis process are: syntactic parse trees, semantic structure graphs, annotated semantic structure graphs (i.e., the CT-Map), 2 1/2 D sketches of the geometry of the scenario (i.e., engineering drawings), a 3D-Map of the scenario geometry, and finally, a synthesized map of the scenario that links these components together i.e., the Three Dimensional Conceptual Thematic Map.

1. INTRODUCTION—A MENTALIST PERSPECTIVE

While it is true that "a picture may be worth a thousand words," it is equally true that a word may in turn evoke a thousand pictures. (Think, for example, of the images associated with the word 'automobile'.) Additionally, it is not only possible to 'see' words and 'talk' about images, but it is also possible to feel, taste, touch, and smell words and images. From observations like these, and the reader could doubtless supply many more, it seems that words and images are as real as anything else in our world. Studying the characteristics of words and images in terms of the information structures that might represent them, may help in understanding and acting on our external world.

As I focus attention on myself, (R.R.), I am conscious of a rush of sensations: the feel of the chair I am sitting on and the stiffness of my back, the smell of pipe smoke in the air in my office, or again the click of the computer keys as I type with the hum of the computer's disk drive in the background, and on and on. We find ourselves routinely in situations such as this, and can comfortably talk about entities that we see, hear, feel, smell and taste, as well as entities best characterized as actions, events, or states. Since we *can* inter-express these modalities, it is

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX logical to assume that there are at least pair-wise links between modality specific representations at some level or levels. From this, an initial simplifying assumption might be that there is just one central level at which all the modality representations are in registration. Additional evidence however, adduced by Marr [1, pp. 35] and Jackendoff [2, pp. 248], suggests there are at least two central levels, Conceptual Structure and 3D Object Structure, autonomous but closely linked. Furthermore, given our ability to discriminate between the sources of our inputs, i.e., sounds entering my ears and sights entering my eyes, it logically follows that there are pathways, at least initially separate, that lead to the levels at which inter-modality communication can take place. As will be elaborated throughout this paper, we propose to take such experiences and their associated lines of reasoning seriously, as examples of psychologically real entities constructed and constrained by internal mental processing.

Recognizing or postulating such aspects of our mental apparatus is one thing, but actually presenting some formal account of its structure is quite another! The little scenario above, with one of the authors at a desk trying to report on his experiences for example, is just the tiniest fraction of what confronts an engineering analyst when he or she begins to try to report to others a specification, design or implementation description or action. The task of conveying the information contained in any real world scenario to another human being is both extremely important and extremely difficult. Fortunately, some progress has been made along these lines, and to that work we turn next.

Background for Constructing Scenario Maps

A promising formalization of mental information structures that encompasses both the linguistic and visual cognitive domains has been proposed by Ray Jackendoff [2]. A fundamental premise of this formalism is Jackendoff's mentalist hypothesis: "Meaning in natural language is an information structure that is mentally encoded by human beings" [2, pp. 122]. A considerable number of contemporary psychologists and cognitive scientists have also adopted variations of the 'mentalist position' as evidenced by the following works: *Modularity of Mind* [3], *Frames of Mind* [4], *Multimind* [5], *The Society of Mind* [6], *Language and Perception* [7]. Visual perception, linguistic processing, memory and so on are viewed as specialized 'faculties' in the mind (see [3]). The same idea is also at the foundation of semantic networks whether or not the network is assumed to reflect the 'real' world directly.

Another assumption that is crucial to our approach to the construction of engineering scenarios is Jackendoff's hypothesis of levels [2, pp. 49]:

1. Each faculty of mind has its own characteristic chain of levels of structure from the lowest (most peripheral) to the highest (most central).
2. These chains intersect at various points.
3. The levels of structure at the intersections of chains are responsible for the interactions among faculties.
4. The central levels at which 'thought' takes place, largely independent of the source of input, are at the intersection of many distinct chains.

Using Jackendoff's work, as well as that of other cognitive scientists such as Marr [1,8], Lerdahl and Jackendoff [9], Sowa [10], Gardner [11], Arbib and Hanson [12], Fodor [13], and Kohler [14], the authors of this paper have developed a working set of hypotheses for scenario construction. We are aiming to *construct* a representation of a scenario in accordance with a version of psychological reality based on the following hypotheses:

The Cognitive Design Hypotheses

A. The human brain/mind may be considered to be a biological information processor.
B. The task of such a processor is to construct/compute descriptions of entities from modality (e.g., linguistic and visual) dependent perspectives having to do with use, purpose, and context. The descriptions consist of information structures of various degrees of detail.
C. It is a psychological reality that people mentally construct intersecting chains of levels of information structures, specific to cognitive modalities, that are composed of primitives

and associated well-formedness rules together with rules of correspondence linking the levels.

D. The mental structures in C are required for an understanding of, and action on, the world.
E. The central levels of information structures (Jackendoff's Conceptual Structure and Marr's 3D-Shape Model) can be refined, extended, and linked.
F. Jackendoff's Conceptual Structures can be refined by annotating them with thematic relations, and they can be extended both by constructing a class of higher level relations that inter-link Conceptual Structures, and by including additional empirically discovered conceptual categories. (These extended and refined representations are what we have called Conceptual Thematic Maps (CT-Maps).)
G. Marr's 3D-Shape Model can be refined by including more shape types as primitives and extended by including multiple types in the visual field. (These extensions are what we have called 3D-Maps.)
H. Linking the CT-Map with the 3D-Map to form one document, the 3D-CT Map, results in additional modeling power.
I. To the extent that an external information structure emulates an internal (mental) information structure, we assume a more accurate transfer of information between human beings will occur.
J. A representation that incorporates many modes or faculties is better than one that incorporates fewer modes.

Figure 1 shows an overview of the levels that need to be considered within the representationalist frame being advanced here. The box shapes in the figure represent levels treated by Jackendoff and Marr, from language and vision, respectively. The diamond shaped objects represent extensions proposed by the authors.

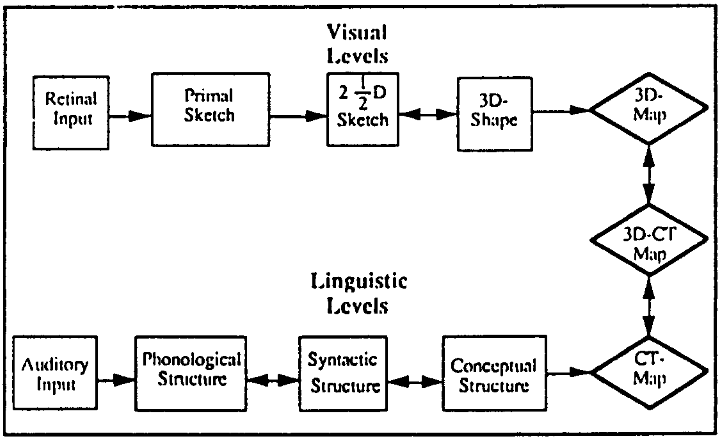

Figure 1. Representationalist levels encompassed by Cognitive Design.

Our hypotheses find further support in the work of Tony Buzan's *Mind Maps* [15], Gabrielle Rico's *clustering* as a writing technique [16], Novak and Gowin's *Propositional Concept Maps* [17], and Betty Edward's *Right Brain Drawing Skills* [18]. In psychology, Neuro-Linguistic Programming therapy has a similar basis [19]. While engineering has always used drawings and text, the balance was, and is, often on the side of text. In the field of systems analysis and requirements specifications, for example, this imbalance of text over graphics fostered the Structured Analysis evolution in the 1970's and 1980's [20]. These methodologies began to accord equal status to both text and graphical notations such as *dataflow* diagrams and *Entity Relationship* boxes. More recently, whole fields of data analysis and statistics are emphasizing the visual mode, with the latest subdiscipline being 'scientific visualization' [21]. In this paper we will briefly describe the representation languages of the different linguistic and visual levels shown in Figure 1 and some of the correspondence rules needed to proceed from one level to another. The ultimate representation of an engineering scenario in our approach combines linguistic and visual information from their central levels into a linked Three Dimensional Conceptual Thematic Map (3D-CT Map).

Jackendoff's Conceptual Structures, refined by thematic relation annotations and extended by higher level linking relations, is the semantic net component of the 3D-CT Map. Marr's 3D-Shape level of visual representation, refined and extended to meet the requirements of an engineering drawing, provides the visual component of the 3D-CT Map.

We illustrate the use of our approach by constructing a 3D-CT Map for a Numerical Control part programming problem. Problems such as these arise in describing, for a machine such as a drill for example, how it is to execute some drilling procedure. Initially, the drilling procedure is written down by the programmer in English, usually supplemented by engineering drawings of the part to be drilled and its geometric environment. From this English language description of what is to be done, together with the engineering drawings, the programmer then translates this into a specialized numerical control computer language program that can effect the actual drill movements. In terms of our model of scenario construction, we translated the English language description of the drilling procedure into a semantic net, the Conceptual Thematic Map. To make this translation, we suggest that it is very helpful to first carry out a syntactic parse in order to understand the general language patterns to be encountered. This level of analysis would not be done every time instructions were to be produced, but only when some new category of instructions were to be implemented. The authors have suggested a procedure for analyzing such patterns in "Path and Location Semantics for Part Programming" [22]. The last part of our illustrative example links the semantic and geometric components into a Three Dimensional-Conceptual Thematic Map.

2. THE LINGUISTIC BASES FOR SCENARIO CONSTRUCTION

Representationalist linguistics postulates a chain of levels of autonomous, mentally encoded information structures for the linguistic and visual faculties. For language understanding, these structures begin with the input acoustic array that is then organized into a level of *phonology*, thence to *syntax* and finally to a *conceptual semantic* structure. What has emerged from the study of these structures is the need to posit primitives specific to each such level together with a set of well-formedness rules that guide the construction of more elaborate objects at that level. Furthermore, the levels are interlinked, or kept in registration, by sets of correspondence rules that have the effect of mapping constituents of one level into constituents of another level. To give a flavor of this work, we present some results and examples from both the syntactic and the semantic levels.

Within the Linguistic Chain: The Syntactic Level

A well-known exposition of the primitives and well-formedness rules for this level has been presented by Jackendoff [23] and Chomsky [24,25]. The primitives of this level consist of *word* and *phrase* categories. For Indo-european languages, the major word (lexical) categories are Verb (V), Noun (N), Adjective (A), Preposition (P), and Adverb (Adv). Expansions of these lexical categories by attached specifiers, complements and adjuncts result into phrasal categories. Top level categories, called *major phrasal categories*, consist of: Sentence (S or V'''), Noun Phrase (NP or N'''), Adjective Phrase (AP or A'''), Preposition Phrase (PP or P'''), and Adverb Phrase (AdvP or Adv''').

These primitives can be combined in accordance with *phrase structure rules*, possibly followed by additional *transformational rules.* We will use the "X-Bar" syntax description method developed by Jackendoff [23] in order to describe syntactic structures. The main rule states that for any given major phrasal category X''', where X is a category variable that stands for any of the lexical categories, the *head* (key constituent) of this phrase is a word category of the same type as X, (see Figure 2). The intermediate categories: X' and X'', along with the other categories, provide a hierarchical structure that is motivated by syntactic and non-syntactic (phonological, morphological, distributional, and semantic) evidence; see Radford [26]. An example of instantiating Figure 2 is

"The operation of drilling by drill A"

As shown in Figure 3 'operation' is the key or head constituent, 'of drilling' attaches at the first level N′ of the syntactic phrase tree, and 'by drill A' attaches at the second level N″. The level of attachment of the latter two phrases reflects their closeness to the key constituent in both syntactic and semantic ways. The notion of closeness is illustrated by changing the order of the two prepositional phrases. From our intuitions about language, this next phrase would be judged as 'not quite right' by most native speakers.

"The operation by drill A of drilling"

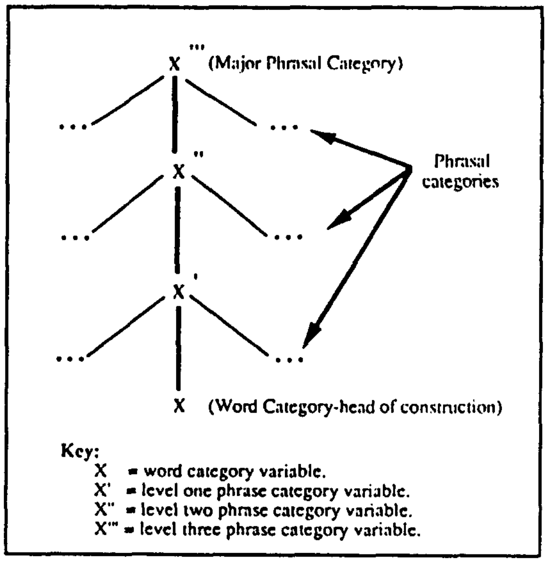

Figure 2. Hierarchical structure of English sentences.

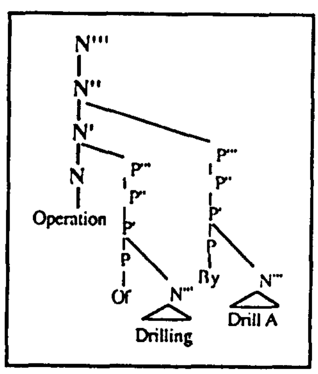

Figure 3. The syntactic structure of "(the) operation of drilling by drill A."

What is important for our work is the connection between the syntax and the semantics wherein the variety of meanings we require for our expressions must be reflected in the form that we express them in. In other words, attachment of phrase categories at various levels in the syntactic phrase trees relate to the types of semantic distinctions we would like to make. Although there is still considerable ferment in the linguistic community over what the fine structure of the syntactic trees look like, their overall shape is widely accepted. The trees exhibit hierarchy and precedence, with a particular phrase being able to trace its major characteristic down to the lexical category at the tip of its subtree.

Within the Linguistic Chain: The Semantic Level

At the autonomous level of semantics, the clearest statement of its structure is again due to the developments of Jackendoff, in his books, *Semantics and Cognition* [27] and *Semantic Structures* [28]. The primitives and well-formedness rules are basically as follows:

1. *Conceptual/Semantic Constituents* represent the primitive units of thought at the conceptual/semantic level. Each such unit of thought has a 'major feature' that determines its categorization as one of a number of *Major Ontological Categories* (MOCs). These MOCs are to be established empirically and Jackendoff uses a linguistically based approach that provides some evidence for various putative categories. The features identified include at least the following: *Event*, *State*, *Path*, *Location* (*or Place*), *Thing* (*or Object*), *Property* (*or Attribute*, *Characteristic*), *Time*, and *Amount* (*or Measure*). There are others yet to be determined, but, these features are sufficient for our work in the location and motion domain. A semantic constituent consists minimally of a (semantic) function, an interpretation of a lexical item (over in the Syntax), that maps into a particular MOC. The semantic function may take arguments that are in turn semantic constituents (nesting).
2. *A formal grammar for Conceptual Semantics*: Given the primitives (semantic constituents) at a level, there is also a need to specify the rules under which these primitives can be decomposed, i.e., *well-formedness rules*. Jackendoff has presented some of these formal *rules of decomposition* which express the *typical* (preference) decomposition of some semantic constituents in the spatial semantic field. We have devised a graphical notation to represent these rules (Figure 4). In this format, semantic constituents are represented by boxes, each box having two basic entries: a semantic function with zero or more arguments and its associated MOC (italicized). Note that the outer box is simply a frame box that is used to border all the figures.

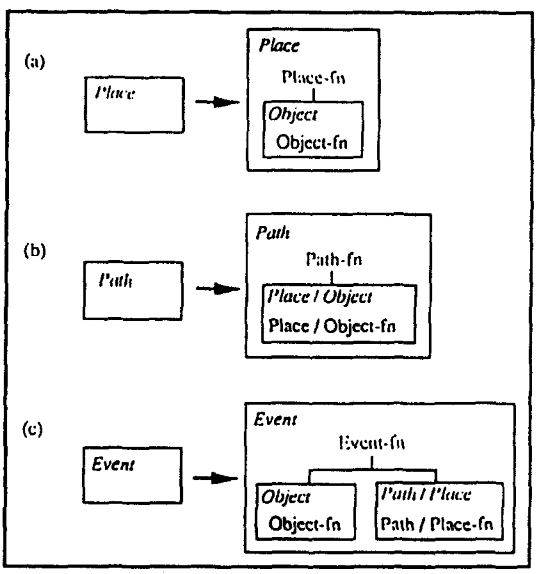

Figure 4. Typical *rules of decomposition* of the semantic constituents 'Place,' 'Path,' and 'Event.'

In the first rule (Figure 4a), the *Place* semantic constituent typically decomposes into a Place-function that has as its argument an *Object* type semantic constituent. For example, 'on the table' has the semantic structure shown in Figure 5.

In the second rule (Figure 4b), the *Path* semantic constituent has two decomposition possibilities. The first has a *Place* semantic constituent as the argument to the Path-function, while the second possibility would have an *Object* type semantic constituent as the argument to the Path-function. Figure 6 shows the semantic structures of 'from under the table' and 'to Spain.'

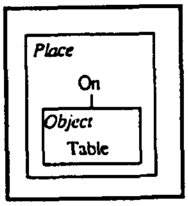

Figure 5. Semantic structure of "on the table."

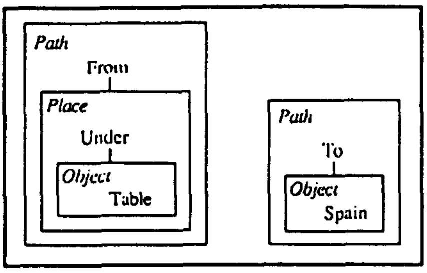

Figure 6. Semantic structure of "from under the table" and "to Spain."

The first structure in Figure 6 is an example of a path that is bounded by a location with respect to an object while the second structure is an example of a path that is bounded by an object. In the last rule (Figure 4c), the *Event* semantic constituent shows that the Event-function requires two arguments for its satisfaction, the first argument is of type *Object* and the second is either of type *Place* or *Path* (the slash '/' indicates alternative choices). The semantic structures in Figure 7 represent 'the tool moves along the plane' and 'Jim sat on the chair.'

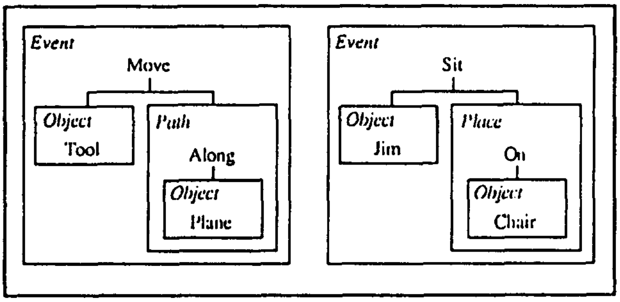

Figure 7. Semantic structure of "the tool moves along the plane" and "Jim sits on the chair."

Additionally, the box (semantic constituent) may contain semantic constituents that are other than functional arguments. In this paper, only one other such type of constituent will be considered and that will be *restrictive modifiers*. Whereas the presence of a functional argument is required in order to define the semantic function corresponding to those lexical heads that require complements or subjects, the purpose of a restrictive modifier is to limit the ways the sentence may be extended. For example, in 'Jim sits on the chair' the semantic function 'sits' requires the interpretation of the two arguments 'Jim' and 'on the chair' for its definition. In the sentence 'Jim sits on the chair by the window,' the prepositional phrase 'by the window' is a restrictive modifier and corresponds to a semantic constituent that is shown in our notation as a free floating box (Figure 8). The semantic purpose of this constituent is to modify the major ontological category, in this case, an *Event*.

Correspondence Rules: From Syntax to Semantics

According to the representationalist linguists, there is a set of correspondence rules that link the syntactic structures to the semantic structures. An important subset of these rules links major

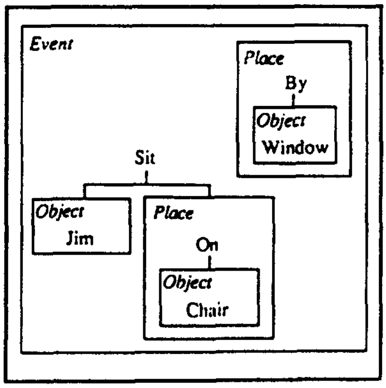

Figure 8. Semantic structure of "Jim sits on the chair by the window."

phrasal categories from the syntax with semantic constituents in semantic structure. The basic correspondence rule of this subset is that each major phrasal constituent, i.e., an X''' syntactic constituent, corresponds to a semantic constituent [27, pp. 67]:

$$X''' \longrightarrow \text{Semantic Constituent}$$

So a NP (N'''), Sentence (V'''), PP (P'''), AP (A'''), and AdvP (Adv'''), corresponds to a semantic constituent in semantic structure. Other correspondence rules provide interpretations of X''' constructions, depending on the level at which these construction attach in the syntactic hierarchy (Figure 9).

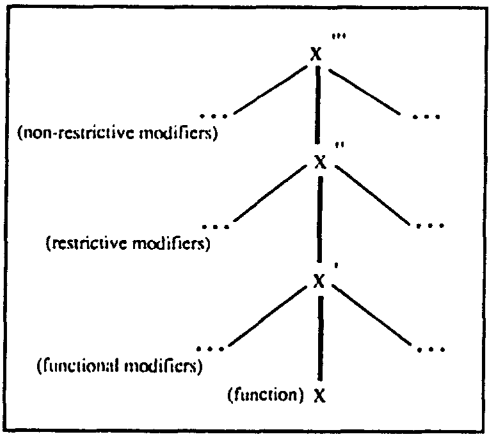

Figure 9. Syntax-Semantic linkage.

In general, phrases that attach at the first level (X' level) map to functional arguments (to a semantic function), those that attach at the second level (X'' level) map to restrictive modifiers, and those that attach at the third level (X''' level) map to another kind of interpretation (called non-restrictive modifiers) that is not of interest to us here. Various exceptional cases are covered extensively in the X-Bar literature.

An Example of Syntax, Semantics and Correspondence

An example of the correspondence between the syntactic level and the conceptual/semantic level is shown by means of Figure 10 and Figure 11. Starting with the English sentence given below, we show its syntactic X-Bar representation in Figure 10 and then show the mapping to the corresponding semantic structure in Figure 11.

"The drill moves from Point 0 (P0) to before Plane 1 (PL1)"

(Omitting the word 'before' would make the statement ambiguous since the word 'to' should reference a path concept rather than a location concept. Since the intent here is to specify that the drill is to end up at a location, we need to specify that requirement explicitly and hence the word 'before' has been inserted.)

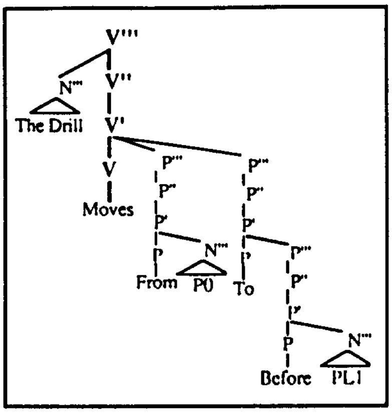

Figure 10. Syntactic structure of "The drill moves from P0 to before PL1." Constituents under the 'triangle' possess a more detailed internal structure that is not relevant to this analysis phase.

The head of the sentence, as shown in Figure 7, is the verb 'move.' The subject, 'The drill,' is an N''' which always links at the third level (V''') of the main 'backbone' of the syntactic tree. The prepositional phrases 'From P0' and 'To before PL1,' each attach at the first level (V' level). Note that in Figure 11 we represent the verb 'move' by means of the semantic function 'GO.' (There is a large class of spatial and motion verbs that map into just a few semantic functions such as GO, BE, and STAY.)

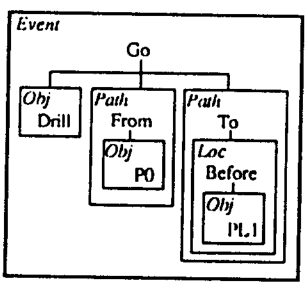

Figure 11. Semantic structure of "The drill moves from P0 to before PL1."

3. THE VISUAL BASES FOR THE SCENARIO MAPS

The work of David Marr on vision understanding, beginning at the MIT Artificial Intelligence Labs in the middle 1970's, in many ways complements and supports the overall mentalist theory advanced by Jackendoff. In fact, Jackendoff uses Marr's work to support his theses [2, Chapter 10]. Marr, in turn, was very aware of the ideas contained in Noam Chomsky's linguistic theory of transformational grammar, which is a theory of *what* the syntactic decomposition of a sentence should be, and not of *how* the decomposition should be carried out. The careful distinction that Chomsky makes between the structure of the syntactic representation on the one hand and the processes that derive these structures on the other hand, is the same distinction that Marr makes in his corresponding theories of vision representation structures versus the processes that manipulate them. Marr's general program of vision research is stated in Marr and Nishihara's 1978 journal article "Representation and recognition of the spatial organization of three-dimensional shapes" [29]:

> "We have studied vision as a process that assembles descriptions in a number of representations, each specializing in some aspect of the visual scene with later ones building on the information made explicit by those before them. This approach is suggested by several experimental findings [30], and is consistent with the *principle of modularity* [31] which states that any large computation should be split up into a collection of small, nearly independent, specialized sub-processes. If visual information processing were not organized in this way, incremental changes in its design would be unable to improve one aspect of the process's performance without simultaneously degrading the operation of many others."

It should be noted here that Marr's work was concerned with describing only a *single* object in the visual field and he left for later researchers the task of extending his work to more complex fields. Marr's work involved discovering to what extent the individual steps of visual processing could be explicated and formalized. How can our brains extract a three dimensional object from what are initially grey scale and color variations in the retinal array? Briefly, what Marr and his colleagues did was to describe a series of levels at which different kinds of information were organized. Each level has its own kinds of primitive objects and well-formedness rules just as in the linguistic chain. Furthermore, Marr was able to hypothesize the algorithms that would enable the organism to derive succeeding levels of information from previous or lower levels, in short, correspondence rules linking visual levels. Thus Marr's work found the same pattern of autonomous representation levels, primitives, rules of well-formedness, and correspondence rules between levels, as did Jackendoff's. Marr comments on the levels of explanation idea [1, pp. 336]:

> "The levels idea is crucial and perception cannot be understood without it—never by thinking just about synaptic vesicles or about neurons and axons, just as flight cannot be understood by studying only feathers. Aerodynamics provides the context in which to properly understand feathers."

Marr's Levels of Visual Representation

Tracing the visual levels of Figure 1, the initial raw *retinal image*, may be thought of as a 2-dimensional array of intensity levels, and the first primitive description of the image is called the *primal sketch*. The primal sketch makes explicit the amount and disposition of the intensity changes in the retinal array. It is hierarchical and consists of primitives at the lowest level representing raw intensity changes and their local geometric structure. The higher level consists of groupings and alignments constructively built up from the lower primitives. Primitives proposed by Marr for the lower sub-level include '*place tokens*' that correspond to oriented edge or boundary segments or to points of discontinuity. *Bars* are proposed to account for (roughly parallel edge pairs) or to their terminations, and *blobs* are hypothesized to account for doubly terminated bars.

The next level is the 2 1/2 D sketch. This is a viewer centered representation of the depth and orientation of the visible surfaces. Primitives at the 2 1/2 D level are the local surface orientation, distance from viewer, discontinuities in depth, and discontinuities in surface orientation. Since this level presents the geometry of the surfaces visible to the viewer, including cues as to depth and orientation of the surfaces, it is more than a flat 2D representation, but since this representation doesn't indicate volumes, it is not yet a 3D model. Another feature of this level is its modularity and its correspondence with the primal sketch. Marr's analysis shows this level to be the convergent target of a number of independent processes, such as stereopsis, surface contours, and shading. These processes take the primal sketch information as their input and return information that is integrated at the 2 1/2 D level. (Jackendoff has proposed several enrichments to Marr's 2 1/2 D sketch that bring out the possibilities of its structural description in a manner closer to the linguistic modeling approach [2, pp. 331].)

It should be noted that many current computerized drawing and drafting packages are, in effect, *2 1/2 D* modelers since they allow only a set of two dimensional views, with additional information appended to each view so that the human viewer can then use these cues to mentally reconstruct a three dimensional object from the presented set. A case in point is the well known CADAM (Computer-Graphics Augmented Design and Manufacturing System) package. In this computer assisted drafting package, the designer represents a 3D object by drawing a number of 2 1/2 D views (e.g., top view, side view, bottom view, isometric view) of the part or work piece, with each drawing being annotated with dimensions and possibly labels. The viewer then looks at a consistent set of these views and mentally reconstructs the original 3D object.

Marr's *3D-Shape* model of a single object in the visual field, is at the end of the visual pathway and is the result of the organisms attempt to deliver an invariant shape description to the central level of visual representation. According to Marr, this representation [1, pp. 37]:

> " ... describes shapes and their spatial organization in an object centered coordinate frame using a modular hierarchical representation that includes volumetric primitives (i.e., primitives that represent the volume of space that a shape occupies) as well as surface primitives."

A key feature of the 3D-Shape model is that it is object centered rather than viewer centered and it is volumetric oriented rather than surface oriented. This allows the object to have a shape and size invariance regardless of the viewer's orientation. This also makes these representations most suitable for encoding into long term memory and thus be available for the tasks of recognition and categorization at a later time. Another crucial feature of this model structure is that it is hierarchical. The 3D structure of an object is not just a monolithic 'figure in the head' that can only be scaled larger or smaller, but rather it is a hierarchical structure that consists of parts and parts within those parts. See Figure 12 for an example of a human figure decomposed under this scheme. From this figure you can see that the coarsest description of the figure is simply a cylinder that is defined by a vertical axis down its center plus a radius. The cylinder can then be decomposed into a body, a head, and four limbs, each of which is a cylinder defined by its own axis and radius. The attachment of the limbs to the body illustrates the idea of an object centered coordinate system wherein the angles that the limbs make with the body are specified in terms of the body axis. Each limb is then decomposed in terms of its own coordinate system. By means of this type of recursive decomposition, once the main axis of an object is specified, the position of the parts are specified automatically.

The correspondence rules that link the 2 1/2 D level to the 3D level are still very much under development but it is the case that some 3D structure can be shown to be derived from 2 1/2 D information [1]. Other principles that would explain how we can construct 3D shape from extremely under-determined 2D surface representation still await adequate explanation. For our purposes it is sufficient to note that there is a correspondence and the critical questions to ask are what is the structure of the information that allows both the correspondence and levels to be manipulated.

Recently developed computer packages directly support several different approaches to three dimensional modeling and one such approach closely matches Marr's underlying assumptions. That approach is embodied in the architectural graphics standard, *Programmer's Hierarchical Interface Graphical Standard* (PHIGS).

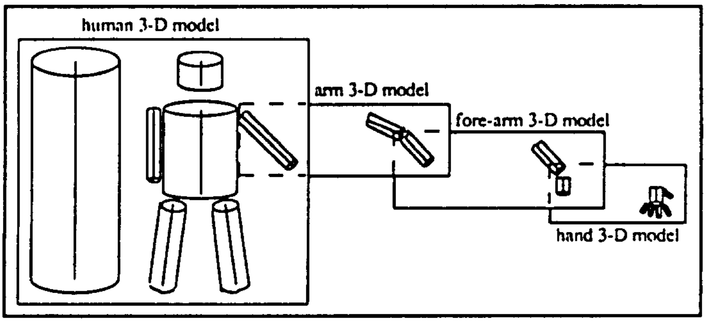

Figure 12. Marr's 3D object centered model (adapted from [1]).

In assessing Marr's work, the cognitive scientist Howard Gardner has stated: "Marr has put forth a plausible account of how an organism may actually proceed from intensities of light to the parsing of the objects in a scene. He has provided a mechanistic account of perception which seems internally coherent and consistent as well with converging evidence drawn from several cognitive disciplines. And even if his particular account turns out to be wrong or incomplete, or if the approach of the parallel processors carries the day, he has defined the likely parameters for future debates about early visual perception" [11, pp. 322].

4. REPRESENTING AN ENGINEERING SCENARIO

When making assumptions and claims about human cognitive capacities, as we are trying to do in this paper, the best way to clarify and make these issues more concrete is to illustrate them in a particular domain. In this section, an example from the field of Numerical Control (NC) part programing is presented. NC part programming is a sub-domain of manufacturing engineering concerned with producing the instructions to machines that will affect the actual cutting, drilling, milling, or other processes that physically alter material in order to produce parts. In this scenario we trace the cognitive design steps that might be undertaken by an analyst as he produces the programming instructions. Roughly, the task of the part programmer in this scenario will be to construct the machine instruction set that will cause a drill to start from a given position, move to a given location, drill a 1 inch hole in a cylindrical workpart, and then return to a particular 'home' location. Figure 13 shows an initial sketch of the environment.

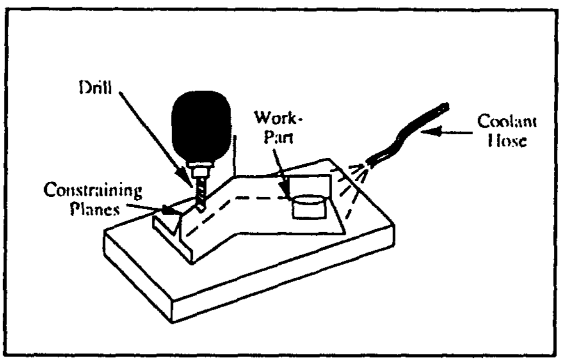

Figure 13. Initial sketch of the scenario environment.

The Engineering Scenario Context: NC Part Programing

At the implementation level, NC part programing is a computer-assisted machining approach in which a computer language controls the movement of the tool with respect to the workpart [32,33]. In NC part programing, the main task of the part programmer is to define the geometry of the workpart and specify the tool path and operation sequence. This task can be defined in terms of four types of statements: (1) geometry definition statements (e.g., points, lines, circles, etc.), (2) tool path instruction statements (e.g., the drill moves from point A to point B), (3) post-processor statements (e.g., the feed rate is 2.7 in./min.), and (4) auxiliary statements (e.g., the cutter diameter is .600 in.) [32]. The computer's job consists of input translation, arithmetic calculations, cutter offset computation, and any post processor operations.

Suppose we consider the task of a part programmer who is faced with the task roughly described previously. Suppose that now the programmer has refined the scenario to the description given below. Furthermore, the decision has been made to translate the English description to a particular low-level machine-tool language, Automatically Programmed Tools (APT). The English description follows, along with comments in parentheses:

The drill moves from Point 0 (P0) to before Plane 1 (PL1),
('before' defines the type of location of the drill with respect to PL1, Fig. 14, label (a))
then the drill moves right along PL1 to before Plane 2 (PL2), (Fig. 14, label (b))
then the drill moves right along PL2 to on Line 1 (L1), (Fig. 14, label (c))
then the coolant turns to on,
while the drill moves 1 in. down,
then the drill moves 1 in. up,
then the drill moves to P0. (drill returns home)

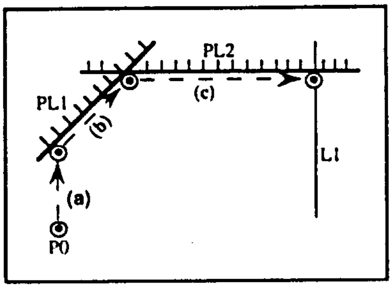

Figure 14. References to the first three English sentences of the scenario description are shown as (a), (b), and (c), respectively.

The APT code which implements this description is:

```
FROM/ PO
GO/ TO, PL1
GORGT/ PL1, TO, PL2
GORGT/ PL2, ON, L1
COOLNT/ ON
GODLTA/ 0, 0, -1.0
GODLTA/ 0, 0, +1.0
GOTO/ PO
```

It should be noted that there are additional statements that would precede these such as defining the coordinates of the Point 'P0' and the location of the Planes 'PL1' and 'PL2'. The first statement in the English description is represented by the first two APT code statements. The rest of the statements correspond one-to-one with the subsequent APT codes. For example, "GORGT/ PL1, TO, PL2" corresponds to the sentence "the drill moves right along PL1 to before Plane 2 (PL2)." In a bit more detail we might interpret this as: "the drill is to move to the right, following Plane 1 (which constrains the drill to follow a path) and terminate at a location just before Plane 2." In English, the word 'To' is ambiguous between its use as a path indicator and as a location indicator. A more careful analysis [27] suggests that it should be used as a path indicator and, if a location is to be indicated, then an additional word, such as 'Before,' needs to be inserted as was done here.

The 2 1/2 D Engineering Drawing and the 3D-Map Component

We would expect the part programing analyst to construct, either implicitly or explicitly, a visual map of the work environment to accompany the English description. When the analyst draws the individual diagrams as shown in Figure 15, he is creating an engineering drawing that is, in effect, a ***2 1/2 D sketch***. Notice that in the individual views, *Top View* and *Front View*, there is no explicit three dimensional component, it is all 2 Dimensional. (The correspondence with

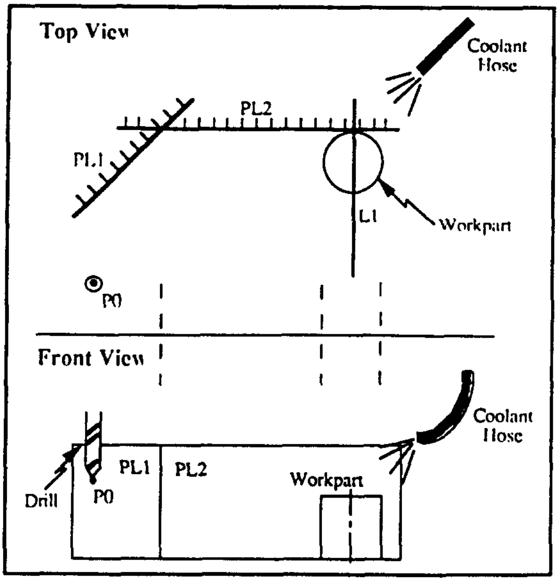

Figure 15. Engineering drawing of Top and Front views of the scenario environment.

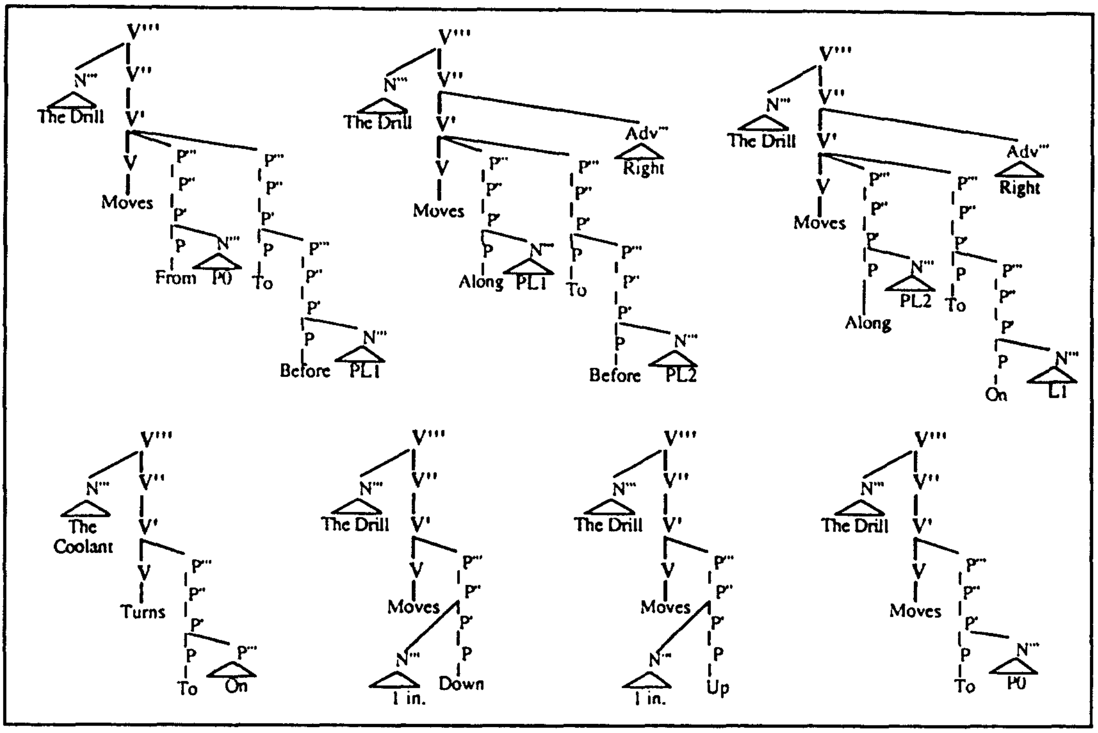

Figure 16. The syntactic structure of the Part Programing scenario.

Marr's 2 1/2 D sketch appears when the programmer annotates the 2D sketches with additional clues as to depths and contour information.) In order for the viewer to realize a three dimensional model from these individual 2 1/2 D drawings, he must mentally synthesize the individual views.

Of course, there are other views that could be drawn (e.g., back, bottom, and isometric), that could be used to illustrate the geometry of the work environment, but for our purpose the drawing shown is adequate. Using this drawing, plus knowing the sequence of tasks desired to be performed as specified in the list of English statements, the part programming analyst translates these specifications into syntactic and semantic models, in accordance with the cognitive design approach.

The Syntactic Component of the Scenario

In our example, each sentence in the scenario text corresponds to one or more code lines in the machine language. The analyst starts the design approach by identifying the internal components of each sentence. This corresponds to implicitly (or explicitly in this case) constructing a syntactic representation of the seven scenario sentences shown in Figure 16.

Although this step in the analysis may seem tedious, our detailed treatment of NC part programming [22] showed that starting with a syntactic analysis helped in identifying the underlying similarities among part programming statements and organizing the vocabulary of the commands. Of course, an experienced programmer who had analyzed such scenarios a number of times might want to skip over this step.

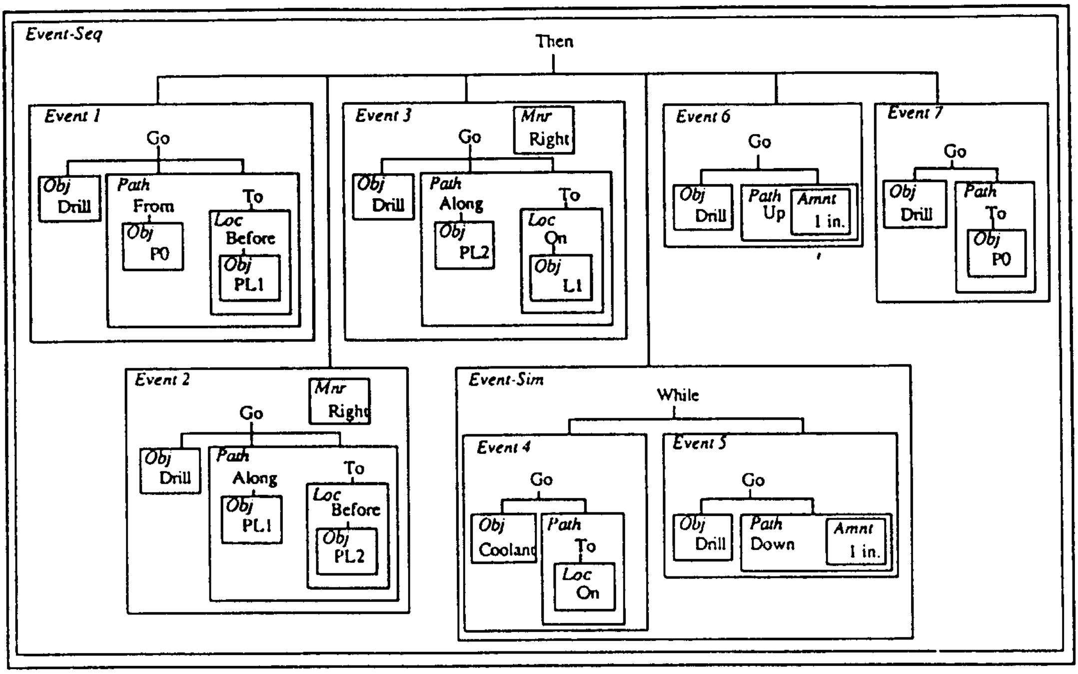

Figure 17. The semantic structure of the Part Programming scenario.

The Conceptual Semantics Component of the Scenario

As the part programing analyst gets ready for the next level of analysis, the *semantic net* level, he starts to categorize constituents of the syntax in semantic terms. He would notice that 'On,' 'Before,' and 'After' are *location* words and 'Along,' 'To,' and 'From' are *path* words. This part of the translation effort relates to the correspondence from syntax to semantics. The semantic structure of Figure 17 illustrates the nesting (hierarchy) of the individual conceptual constituents and also illustrates higher level *communication relations.* The communication relations reflect connections between sentences that are lexicalized by words such as 'then' and 'while.' The nature of these higher level communication relations has not been analyzed in detail by the authors, but we have pointed out work that has been done by Larson [34] that suggests how they might fit into the scheme advanced here. We have thus translated the lexical items 'then' and 'while' as high level semantic functions whose arguments are entire semantic structures. In Figure 17 the semantic function 'then' has event constituents one through seven as its arguments. The semantic function 'while' has event constituents four and five as its arguments.

Thematic Relations for the Scenario

Next, the analyst assigns roles, called *thematic relations,* to selected semantic constituents of Figure 17. The result of these assignments (the underlying theory is detailed in the next section) are the Conceptual Thematic Maps of Figures 18 and 19, shown just for events one, two, four,

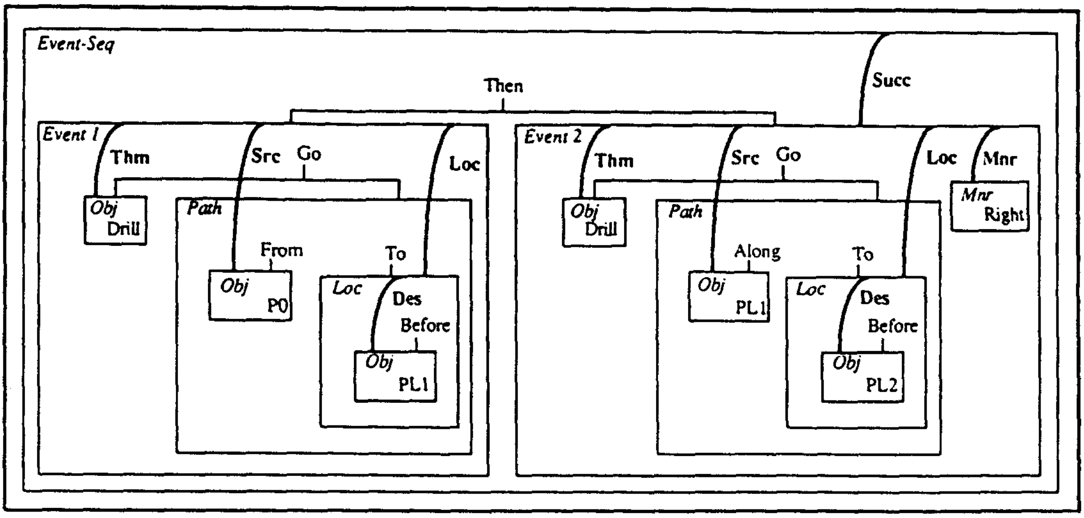

Figure 18. CT-Map of Events 1 and 2.

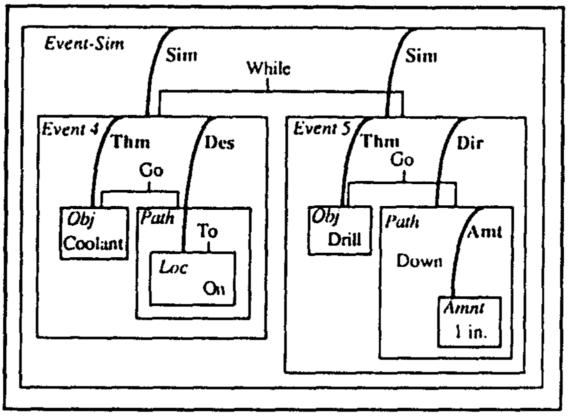

Figure 19. CT-Map of Events 4 and 5.

and five. The purpose of assigning roles to the semantic constituents in these events is to aid the analyst in understanding what commands are actually needed before he is faced with the task of figuring out how they can be expressed in the particular syntax of the machine language. A further reason for performing this thematic role analysis relates to disambiguation tasks of the type treated earlier with respect to the use of the word 'To.' This was a case in which the APT code treated it as a location indicator when it should actually be a path indicator.

To see how the thematic analysis might help the analyst, consider the roles indicated as **theme** shown in Figure 18 and 19. These themes explicitly indicate the object being moved or affected. In one case, the **theme** is the 'Drill,' and in the case of Figure 19, Event 4, the **theme** is the 'Coolant.' It is a matter of some interest that the machine code does not express these themes in the same way in the syntax. In the case of the *object* 'Drill,' for example, it is not even indicated in the code, its presence simply being assumed. On the other hand, the **theme** role of the 'Coolant' in Figure 19 Event 4, is explicitly recognized by the APT language. So we note in this language that when the **theme** is explicitly recognized by the APT code, it is translated to an APT command placed at the start of a geometry or a postprocessor statement. When it is not recognized, the analyst is confronted with the task of keeping track of this hidden information in some manner, perhaps with the CT-Map. The APT code below explicitly shows the syntax for the themes COOLaNT, FEeDRATe, and POINT.

```
COOLNT/ON           (the coolant turns on)
FEDRAT/2.5          (the feed rate is 2.5 in./min.)
P0 = POINT/0,0,0    (point 0 is at (0,0,0)).
```

General Characteristics of Thematic Relations

The position taken in this paper with regard to thematic relations is that they are labels for particular semantic structural configurations, and more specifically, the pairwise relation of conceptual/semantic constituents. Some of the role labels are application independent, such as the **theme** role, while others may be chosen based on the application such as the **location** role in the part programming scenario. Given this perspective, some semantic constituents may have more than one role; for example, 'product A' in 'inspect product A' has two roles, **theme** and **patient**. Due to the possibility of multiple roles for a semantic constituent, plus other arguments, Jackendoff has claimed that thematic roles are not primitives in the semantic model, [35, pp. 378]:

> "Thematic Relations are to be reduced to structural configurations in conceptual structure: the names for them are just convenient mnemonics for particularly prominent configurations."

This means that we may pragmatically define thematic relations between connected concepts based on the purpose of the configuration (structure) in which these concepts participate. Some of the more well known thematic relations are:

Theme: the object in motion or being located.

Source: the object from which motion proceeds.

Goal: the object to which motion proceeds.

Agent: the first argument of the event function CAUSE(i, j).

John Sowa [10], on the other hand, uses the same thematic relation mnemonics but with descriptions that are suitable for use in Conceptual Graphs. We notice that the last three thematic relations described by Jackendoff are also part of Sowa's set of Conceptual Relations [10]. Sowa, however, uses the name **Destination** or **Location** (depending on use type) for **Goal**. So, the entries for Sowa's Conceptual Relations are not definitions, rather they are constraints on their use in Conceptual Graphs. In other words, the interpretation of these relations may vary depending on their use, as was evident in Jackendoff's somewhat differing interpretation above. Some of Sowa's thematic relations are:

Source: links an act to an entity, from which it originates.

Destination: links an act to an entity, toward which the action is directed.

Agent: links an act to an animate where the animate concept represents the actor of the action.

Manner: links an act to an attribute.

Sowa, like Jackendoff, treats conceptual (thematic) relations as convenient linkages between concepts (but see the distinction between concepts and conceptual/semantic constituents treated in the next section). Sowa has also provided a mechanism for defining other relations as the need arises.

Other descriptions of conventional thematic relations are provided by Radford [26]:

Theme (or **Patient**): entity undergoing the effect of some action.

Source: entity from which something moves.

Goal: entity towards which something moves.

Agent (or **Actor**): instigator of some action.

The Conceptual Thematic Map

Given these views on the nature of Thematic Relations, the analyst is prepared to assign thematic roles to semantic constituents in a semantic structure such as Figure 17. The resultant representation shown in Figures 18 and 19, is what we have called the Conceptual Thematic Map (CT-Map) and is a type of semantic net. We have shown these relation labels as linking their respective semantic constituents. In Figure 18, the **Theme (Thm)** is linked to the *Event 1* semantic constituent by a bold labeled line. Note that the roles are read from the 'inside out.' This means that the role indicates how a more deeply nested constituent relates to an outer, containing constituent. The *object* 'Drill,' for example, plays the role of **theme** within the semantic constituent whose category is that of an *event*. The *object* 'P0' plays the role of **Source (Src)** with respect to the semantic constituent of category constituent *event*. Looking deeper within the nested constituents of *Event 1*, we see that the constituent of category *Object* (*Obj*), containing 'PL1,' plays the role of a **Destination (Des)** with respect to its 'parent' concept of category *Location* (*Loc*). The reader may also note that not every constituent need be assigned a role within the overall *Event 1* constituent. This is the case with the *path* constituent since there was no need in this particular analysis to assign it a role. The thematic role assignments are thus dependent on the semantic structure and are pragmatically assigned as needed. Another interesting role assignment in Figure 18 is shown by the label **Succ (Successor)**. This is a role that indicates the connection between the constituent *Event 2* and the overall constituent *Event-Sequence*. This role may be interpreted as showing a temporal connection between *Event 1* and *Event 2*. Figure 19 illustrates further thematic role assignments involving **Amount (Amt)** and **Direction (Dir)**.

It is a valuable exercise to briefly note points of contact between John Sowa's Conceptual Graphs and our CT-Maps, since Sowa's graphs are machine processable and have direct translations to predicate logic. In particular, we have analyzed some of the points of comparison with respect to the overall structure of the two representations and their respective interpretation of conceptual relations and thematic roles. For comparison purposes, the *Event 1* constituent of Figure 19 is re-drawn in Figure 20 together with an 'equivalent' Conceptual Graph. Note that Sowa uses the term 'conceptual relation' to refer approximately to what we are calling thematic roles. An important distinction between the two approaches is that Sowa's linkages are between *concepts*, while the Conceptual Thematic Map links are between *conceptual/semantic constituents*. In both cases, the idea is to state some constraint between concepts/conceptual constituents where the constraint is either pragmatically asserted or structurally derived. In the Conceptual Graph of Figure 20, Sowa would consider 'P0' and 'Go' to be concepts related by the conceptual relation **Source**. In the CT-Map the interpretation of linkage is a little more involved. Here, we consider 'P0' to be a zero argument semantic function that generates its conceptual constituent of category *object*. This constituent is linked to the overall conceptual constituent *Event 1* by the thematic relation **Source**. Thus these connections are between conceptual constituents. There is a way though to translate between the two viewpoints. We could let the semantic function 'P0' stand for its conceptual constituent and further interpret 'P0' itself as a concept. Similarly, although 'Go' is a semantic function that generates the conceptual constituent *Event 1* when its arguments are filled in, we could consider 'Go' itself to stand for its constituent and further treat it as a concept. Under this interpretation, we can begin to see a way to translate between the two representations. These considerations led to the development of the Conceptual Graph from the CT-Map as shown in Figure 20.

The 3 Dimensional Conceptual Thematic Map

In this section, we describe a procedure for constructing linked visual and linguistic maps in the spatial domain. The intent here is to construct a document, or documents, that reflect the theoretical hypothesis that there are two central levels of mental information structures, the three dimensional object representation structure (i.e., 3D-Map), and the semantic conceptual structure annotated with thematic relations (i.e., the CT-Map/semantic net). Furthermore, these information structures are linked by correspondence rules that keep structural fragments at one level in registration with certain fragments at another level. We propose to emulate

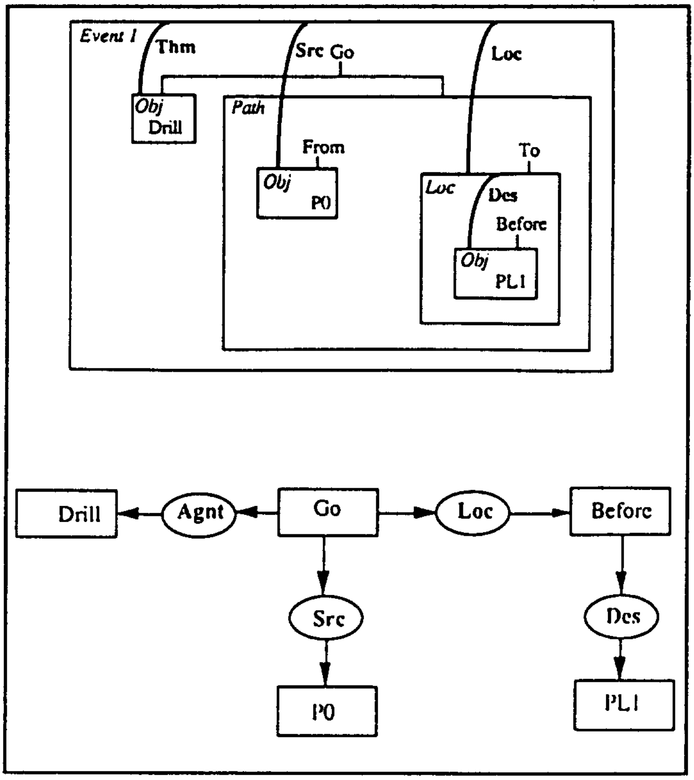

Figure 20. The CT-Map and CG of Event 1 in the scenario example.

this hypothesis externally by constructing three dimensional diagrams, semantic net structures, and interpreting the correspondence rules as the *juxtaposition of these two structures at selected points*. By juxtaposing the three dimensional representation with the linguistic representation on one document, we hope to trigger both modalities into synthesizing a richer representation at both the 3D object level and the linguistic level, as well as strengthening the correspondence links between them.

We start with an intermediate level map that we call a *2 1/2 dimension CT-Map* as shown in Figure 21. The first thing to notice about this map is its underlying 2 1/2 D geometric basis. The reader has already encountered this diagram before in Figure 15, where we discussed the 2 1/2 D sketch as being an information structure at one level of the visual pathway hierarchy. This geometric view of the drilling machine's path and auxiliary objects is now linked with components from the associated CT-Map that was derived from the English language description. For example, the label 'EV1' corresponds to the conceptual constituent, *Event 1* of Figure 17. The meaning of this constituent in turn reflects the original English sentence "The drill moves from Point 0 (P0) to before Plane 1 (PL1)." The label 'EV4' refers to the constituent *Event 4* of Figure 17 and may be traced back to the sentence of the scenario that reads "The coolant turns to on." The other annotations are interpreted similarly.

In Figure 22, we present the Three Dimensional Conceptual Thematic Map (3D-CT Map) that is the culmination of the earlier maps. This final map is built from the 2 1/2 D-CT map described above. First of all, this final map is intended to suggest the multi-modal nature of the scenario that describes a process. The fact that the drilling operation is actually carried out in three dimensions is a feature that we wish to emphasize in the map, and so we have tried to show its position in space. Furthermore, another dimension of the scenario that we wished to capture was its connection to the semantic net that underlies the original English natural language description. To capture this last connection, we place tags on the 3D-CT Map that refer to semantic constituents in the Conceptual Thematic Map. The explanation of the EV1, EV2, ..., EV7 tags are the same as for the 2 1/2 D-CT Map.

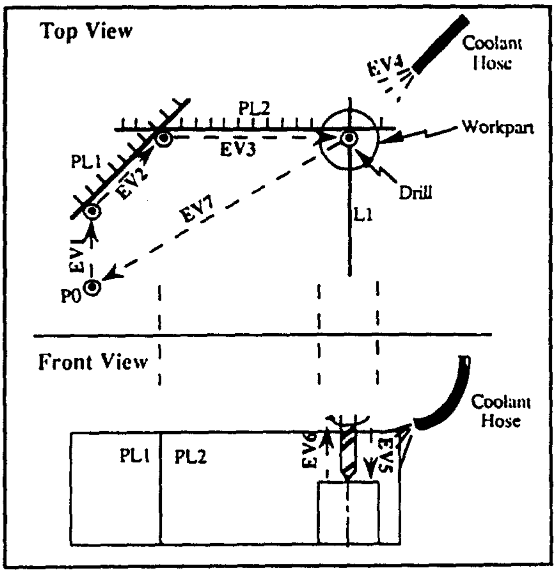

Figure 21. The 2 1/2 dimension CT-Map of the Part Programming scenario.

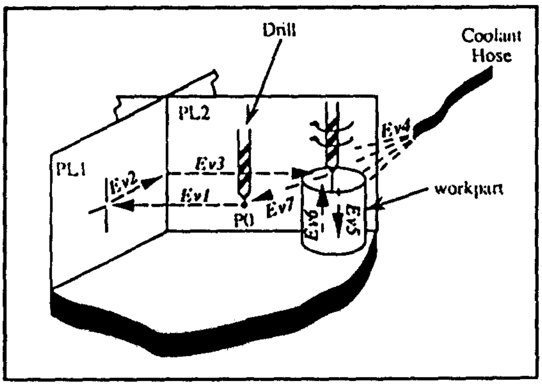

Figure 22. The 3D-CT Map of the scenario example.

The part programmer analyst can now use the 3D-CT Map to *view* the position of a given operation (event) in reference to the environmental geometry (3D-Map) and to other operations, while at the same time he can use the labels on the dashed arrows in the 3D-CT Map (EV1, EV2, etc.) as reference links to the *linguistically*-based CT-Map.

Other researchers have recently begun to try to integrate the linguistic and visual modalities in order to interpret scenarios. In their paper "How near is too far? Talking about visual images," Zernick and Vivier [36] describe their efforts to construct a computational model that uses language clues in order to direct visual processing. Their objective is to detect specified objects in a visually rendered scene. Their approach is to employ a particular type of linguistic processing, 'directive semantics,' that will aid in this search for salient visual objects. Their approach receives both textual and visual input and the analyzed text is used to guide the search. In the same spirit of using the linguistic analysis to direct and constrain the visual search, Srihari and Rapaport in their paper "Extracting visual information from text: using captions to label human faces in newspaper photographs," propose translating the natural language figure caption descriptions of photographs into a semantic net that encodes knowledge about the photographs [37]. They then use this knowledge to draw inferences about the location of visual objects by reference to corresponding inferences drawn on the semantic net.

The correspondence of these works with the author's approach rests on the shared belief that the linguistic and visual descriptions can both be used to constrain the other, thus resulting in a richer synthesized description for the human being. The machine processable objectives of the studies referenced above is however oriented in a somewhat different direction than ours. We are most concerned with seeing to what extent a particular theory of mental representation may be used as a set of design guidelines that would aid in engineering scenario analysis.

5. SUMMARY/CONCLUSION

We have tried to show the general outlines and the possibilities of application of a design approach for constructing engineering scenario maps that is founded on a set of psychological principles articulated most eloquently and coherently by Ray Jackendoff [2] and David Marr [1]. By adopting Jackendoff's mentalist postulate and the hypothesis of levels of explanation of both Marr and Jackendoff, we are led to look for, and explicitly represent hypothesized information structures that have the characteristics of: *autonomous levels*, specified *primitives at a level, well-formedness rules for the primitives at a level*, and *sets of correspondence rules linking levels*. We have summarized our interpretation of Jackendoff's work and others in a set of hypotheses which we called the Cognitive Design Hypotheses. We have used these hypotheses as design guidelines for our scenario development analysis. Our Three Dimensional Conceptual Thematic Map (3D-CT Map) expresses our interpretation of the linked visual and linguistic representation structures that form the basis for human understanding of the engineering scenario under consideration. This 3D-CT Map is composed of a semantic net component and a three dimensional drawing component. The semantic net component is what we have called a Conceptual Thematic Map and is an extension of Jackendoff's semantic structure wherein the semantic constituents are linked by thematic roles. The three dimensional map component is a three dimensional engineering drawing that traces its bases to David Marr's three dimensional shape model that now encompasses the full engineering visual field.

Working through an example scenario drawn from the field of Numerical Control Part Programming statement analysis, our derived 3D-CT map shows essentially a 3D engineering drawing with cross reference information to 2 1/2D engineering drawings, and cross references to a semantic structure augmented by thematic roles that represents the linguistic component of the scenario. The explicit underlying structure of the English sentences allows the analyst to appreciate the organization of the various spatial concepts in anticipation of translating these concepts to a machine language implementation. Additionally, linking these semantic components to a 3D drawing helps the analyst to visualize and verbalize 'what is happening to what, and where,' all on one document.

REFERENCES

1. D. Marr, *Vision*, W.H. Freeman, New York, (1982).
2. R. Jackendoff, *Consciousness and the Computational Mind*, The MIT Press, Cambridge, (1987).
3. J.A. Fodor, *The Modularity of Mind: An Essay on Faculty Psychology*, The MIT Press, Cambridge, (1983).
4. H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books Inc. Publishers, New York, (1983).
5. R. Ornstein, *Multimind: A New Way of Looking at Human Behavior*, Boston, (1986).
6. M. Minsky, *The Society of Mind*, Simon & Schuster, New York, (1986).
7. G.A. Miller and P.N. Johnson-Laird, *Language and Perception*, Harvard University Press, Cambridge, (1976).
8. D. Marr and E. Hildreth, Theory of edge detection, *Proc. R. Soc. Lond. B 207*, 187–217 (1980).
9. F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*, The MIT Press, Cambridge, (1983).
10. J. Sowa, *Conceptual Structures: Information Processing in Mind and Machines*, Addison-Wesley Publishing Company Inc., (1984).
11. H. Gardner, *The Mind's New Science: A History of the Cognitive Revolution*, Basic Books Inc. Publishers, New York, (1985).
12. M.A. Arbib and A.R. Hanson, Vision, brain, and cooperative computation: An overview, In *Vision, Brain, and Cooperative Computation* (Edited by M. Arbib and A. Hanson), The MIT Press, Cambridge, (1987).
13. J.A. Fodor, Mental presentation: An introduction, In *Scientific Inquiry in Philosophical Perspective*, (Edited by N. Rescher), pp. 105–128, University Press of America, New York, (1987).
14. W. Kohler, *The Task of Gestalt Psychology*, Princeton University Press, Princeton, New Jersey, (1969).
15. T. Buzan, *Use Both Sides of Your Brain*, E.P. Dutton, New York, (1983).
16. G.L. Rico, *Writing the Natural Way: Using Right-Brain Techniques to Release Your Expressive Powers*, Tarcher Inc., Los Angeles, (1983).

17. J.D. Novak and D.B. Gowin, *Learning How to Learn*, Cambridge University Press, Cambridge, (1984).
18. B. Edwards, *Drawing on the Right Side of the Brain*, Tarcher Inc., Los Angeles, (1989).
19. R.B. Dilts, *Roots of Neuro-Linguistic Programming*, Meta Publications, California, (1983).
20. T. De Marco, *Structured Analysis and System Specification*, Yourdon Inc., New York, (1979).
21. B. McCormick, Advanced visualization environments: Knowledge-based image modeling, In *Visualization in Supercomputing* (Edited by R. Mendez), Springer-Verlag, New York, (1990).
22. T. Aldowaisan and R. Rucker, A linguistic analysis of part programming statements, *Journal of Manufacturing Systems* 10 (4) (1991).
23. R. Jackendoff, *X Syntax: A Study of Phrase Structure*, The MIT Press, Cambridge, (1977).
24. N. Chomsky, *Studies on Semantics in Generative Grammar*, Mouton, The Hague, (1972).
25. N. Chomsky, *Rules and Representations*, Columbia University Press, New York, (1980).
26. A. Radford, *Transformational Grammar: A First Course*, Cambridge University Press, (1988).
27. R. Jackendoff, *Semantics and Cognition*, The MIT Press, Cambridge, (1983).
28. R. Jackendoff, *Semantic Structures*, The MIT Press, Cambridge, (1990).
29. D. Marr and H.K. Nishihara, Representation and recognition of the spatial organization of three-dimensional shapes, *Proc. R. Soc. Lond.* **B 200**, 269–294 (1978).
30. D. Hubel, and T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol. London* **160**, 106–154 (1962).
31. D. Marr, Early processing of visual information, *Phil. Trans. R. Soc. Lond.* **B 275**, 483–524.
32. M.P. Groover, *Automation, Production Systems, and Computer Integrated Manufacturing*, Prentice-Hall, Inc., New Jersey, (1987).
33. I.H. Kral, *Numerical Control Programming in APT*, Prentice-Hall, (1986).
34. M.L. Larson, *Meaning-Based Translation: A Guide to Cross-Language Equivalence*, University Press of America, New York, (1984).
35. R. Jackendoff, The status of thematic relations in linguistic theory, *Linguistic Inquiry* 1 (3), 369–411 (Summer 1987).
36. U. Zernick and B. Vivier, How near is too far? Talking about visual images, *Proceedings of the 10th Annual Conference of the Cognitive Science Society, Montreal, 1988*, pp. 202–208, Lawrence Erlbaum Associates.
37. R. Srihari and W. Rapaport, Extracting visual information from text: Using captions to label human faces in newspaper photographs, In *Proceedings of the 10th Annual Conference of the Cognitive Science Society, Montreal 1988*, pp. 209–215, Lawrence Erlbaum Associates.